IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 2, Numărul 1, Martie 2023

Procesarea Big Data

Nicolae Sfetcu

Sfetcu, Nicolae (2023), Procesarea Big Data, *IT & C*, **2**:1, 18-22, DOI: 10.58679/IT91785, https://www.internetmobile.ro/procesarea-big-data/

Publicat online: 04.01.2023

Procesarea Big Data

Nicolae Sfetcu
nicolae@sfetcu.com

Big Data Processing

Abstract

The data must be processed with advanced collection and analysis tools, based on predetermined algorithms, in order to obtain relevant information. Algorithms must also take into account aspects invisible to direct perceptions. Big Data in government processes increase cost efficiency, productivity and innovation. Civil registries are a source for Big Data. The processed data helps in critical development areas such as health care, employment, economic productivity, crime, security and natural disaster and resource management.

Keywords: big data

Rezumat

Datele trebuie procesate cu instrumente avansate de colectare și analiză, pe baza unor algoritmi prestabiliți, pentru a putea obține informații relevante. Algoritmii trebuie să ia în considerare și aspecte invizibile pentru percepțiile directe. Big Data în procesele guvernamentale cresc eficiența costurilor, productivitatea și inovația. Registrele civile sunt o sursă pentru Big Data. Datele prelucrate ajută în domenii critice de dezvoltare, cum ar fi îngrijirea sănătății, ocuparea forței de muncă, productivitatea economică, criminalitatea, securitatea și gestionarea dezastrelor naturale și a resurselor.

Cuvinte cheie: big data

IT & C, Volumul 2, Numărul 1, Martie 2023, pp. 18-22
ISSN 2821 - 8469, ISSN – L 2821 - 8469
URL: https://www.internetmobile.ro/procesarea-big-data/

Datele trebuie procesate cu instrumente avansate de colectare și analiză, pe baza unor algoritmi prestabiliți, pentru a putea obține informații relevante. Algoritmii trebuie să ia în considerare și aspecte invizibile pentru percepțiile directe.

În 2004 Google a publicat o lucrare despre un proces numit MapReduce care oferă un model de procesare paralelă. (Dean and Ghemawat 2004) De asemenea, MIKE2.0 este o aplicație în sursă deschisă pentru managementului informațiilor. (MIKE2.0 2019) Mai multe studii din 2012 au arătat că arhitectura optimă pentru a aborda problemele din Big Data sunt cele cu mai multe straturi. O arhitectură paralelă distribuită distribuie date pe mai multe servere (medii de execuție paralelă) putându-se îmbunătăți astfel dramatic vitezele de procesare a datelor.

Conform unui raport al Institutului Global McKinsey din 2011, principalele componente și ecosisteme ale Big Data sunt: (Manyika et al. 2011) tehnici de analiză a datelor (învățarea mașinilor, prelucrarea limbajului natural, etc.), tehnologii de mari dimensiuni (business intelligence, cloud computing, baze de date), și vizualizări (diagrame, grafice, alte afișări ale datelor).

Big Data furnizează informații în timp real sau aproape real, evitându-se astfel latența ori de câte ori este posibil.

Aplicații

Big Data în procesele guvernamentale cresc eficiența costurilor, productivitatea și inovația. Registrele civile sunt o sursă pentru Big Data. Datele prelucrate ajută în domenii critice de dezvoltare, cum ar fi îngrijirea sănătății, ocuparea forței de muncă, productivitatea economică, criminalitatea, securitatea și gestionarea dezastrelor naturale și a resurselor. (Kvochko 2012)

De asemenea, Big Data oferă o infrastructură care este permite evidențierea incertitudinilor, a performanței, și disponibilitatea componentelor. Tendințele și predicțiile în industrie necesită o cantitate mare de date și instrumente avansate de predicție.

Big Data contribuie la îmbunătățirea asistenței medicale prin furnizarea de medicamente personalizate și analize prescriptive, intervenții clinice cu evaluarea riscurilor și analize predictive, etc. Nivelul datelor generate în sistemele de sănătate este foarte mare. Dar există o problemă presantă cu generare de "date murdare", care cresc odată cu creșterea volumului de date, mai ales că cele mai multe sunt nestructurate și greu de utilizat. Utilizarea Big Data în domeniul asistenței medicale a generat provocări etice semnificative, cu implicații asupra drepturilor individuale, viața privată și autonomia, transparența și încrederea.

În media și publicitate, pentru Big Data se folosesc numeroase puncte de informare despre milioane de persoane, pentru a servi sau transmite mesaje sau conținuturi personalizate.

În domeniul asigurărilor de sănătate se colectează date despre "factorii determinanți ai sănătății", care ajută la elaborarea de previziuni privind costurile de sănătate și identificarea problemele de sănătate ale clienților. Această utilizare este controversată, datorită discriminării clienților cu probleme de sănătate. (Allen 2018)

Big Data și tehnologia informației se complementează reciproc, ajutând împreună la dezvoltarea Internetului Lucrurilor (Internet of Things, IoT) pentru interconectarea dispozitivelor inteligente și colectarea datelor senzoriale utilizate în diferite domenii.

În sport, Big Data poate ajuta la îmbunătăți pregătirii și înțelegerea concurenților utilizând senzori specifici, și se poate prezice performanța viitoare a sportivilor. Senzorii atașați mașinilor din Formula 1 colectează, printre altele, date din presiunea în anvelope pentru a eficientiza arderea combustibilului.

În cercetare

În știință, sistemele Big Data sunt folosite intens în acceleratoarele de particule de la CERN (150 de milioane de senzori transmit date de 40 de milioane de ori pe secundă, pentru cca 600 de milioane de coliziuni pe secundă, din care se utilizează după filtrare doar 0,001% din totalul datelor obținute), (Brumfiel 2011) în telescoapele radio astrofizice construite din mii de antene, decodificarea genomului uman (inițial a durat câțiva ani, cu Big Data se poate realiza în mai puțin de o zi), studii climatice, etc.

Marile firme IT utilizează depozite de date de ordinul zecilor de petabyte pentru căutare, recomandări și merchandising. Cele mai multe date sunt colectate de Facebook, cu peste 2 miliarde

de utilizatori activi lunar, (Constine 2017) și Google cu peste 100 de miliarde de căutări pe lună. (Sullivan 2015)

În cercetare se folosește mult căutarea criptată și formarea clusterelor în Big Data. Țările dezvoltate investesc enorm în prezent pentru cercetare în Big Data. În cadrul Uniunii Europene, aceste cercetări sunt înglobate în programul-cadrul Orizont 2020. (European Commission 2019)

Adesea, programele de cercetare folosesc resursele API de la Google și Twitter pentru a obține acces la sistemele lor Big Data, gratuit sau contra cost.

Seturile mari de date vin cu provocări algoritmice care anterior nu existau, fiind imperios necesar să se schimbe în mod fundamental modalitățile de procesare. Pentru aceasta s-au creat ateliere speciale de lucru care reunesc oameni de știință, statisticieni, matematicieni și practicieni pentru a discuta despre provocările algoritmice ale Big Data.

Bibliografie

Allen, Marshall. 2018. "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates." Text/html. ProPublica. July 17, 2018. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

Brumfiel, Geoff. 2011. "High-Energy Physics: Down the Petabyte Highway." *Nature* 469 (7330): 282–83. https://doi.org/10.1038/469282a.

Constine, Josh. 2017. "Facebook Now Has 2 Billion Monthly Users… and Responsibility." *TechCrunch* (blog). 2017. http://social.techcrunch.com/2017/06/27/facebook-2-billion-users/.

Dean, Jeffrey, and Sanjay Ghemawat. 2004. "MapReduce: Simplified Data Processing on Large Clusters." http://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf.

European Commission. 2019. "Horizon 2020." Text. Horizon 2020 - European Commission. 2019. https://ec.europa.eu/programmes/horizon2020/en.

Kvochko, Elena. 2012. "Four Ways to Talk About Big Data." Text. Information and Communications for Development. December 4, 2012. http://blogs.worldbank.org/ic4d/four-ways-to-talk-about-big-data.

Manyika, James, Michael Chui, Jaques Bughin, and Brad Brown. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity." 2011. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

MIKE2.0. 2019. "Big Data Solution Offering - MIKE2.0, the Open Source Methodology for Information Development." 2019. http://mike2.openmethodology.org/wiki/Big_Data_Solution_Offering.

Sullivan, Danny. 2015. "Google Still Doing At Least 1 Trillion Searches Per Year." Search Engine Land. January 16, 2015. https://searchengineland.com/google-1-trillion-searches-per-year-212940.